

Abstract for 2015 iCAS Symposium  
September 13-17, 2015  
Annecy, France

***High Resolution Nature Runs and the Big Data Challenge***

W. Phillip Webster, PhD and Daniel Q. Duffy, PhD  
NASA Goddard Space Flight Center  
Greenbelt, MD USA

NASA's Global Modeling and Assimilation Office at Goddard Space Flight Center is undertaking a series of very computationally intensive Nature Runs and a downscaled reanalysis. The nature runs use the GEOS-5 as an Atmospheric General Circulation Model (AGCM) while the reanalysis uses the GEOS-5 in Data Assimilation mode. This paper will present computational challenges from three runs, two of which are AGCM and one is downscaled reanalysis using the full DAS.

The nature runs will be completed at two surface grid resolutions, 7 and 3 kilometers and 72 vertical levels. The 7 km run spanned 2 years (2005-2006) and produced 4 PB of data while the 3 km run will span one year and generate 4 PB of data. The downscaled reanalysis (MERRA-II Modern-Era Reanalysis for Research and Applications) will cover 15 years and generate 1 PB of data.

Our efforts to address the big data challenges of climate science, we are moving toward a notion of Climate Analytics-as-a-Service (CAaaS), a specialization of the concept of business process-as-a-service that is an evolving extension of IaaS, PaaS, and SaaS enabled by cloud computing. In this presentation, we will describe two projects that demonstrate this shift. *MERRA Analytic Services* (MERRA/AS) is an example of cloud-enabled CAaaS. MERRA/AS enables MapReduce analytics over MERRA reanalysis data collection by bringing together the high-performance computing, scalable data management, and a domain-specific climate data services API. NASA's *High-Performance Science Cloud* (HPSC) is an example of the type of compute-storage fabric required to support CAaaS. The HPSC comprises a high speed Infiniband network, high performance file systems and object storage, and a virtual system environments specific for data intensive, science applications. These technologies are providing a new tier in the data and analytic services stack that helps connect earthbound, enterprise-level data and computational resources to new customers and new mobility-driven applications and modes of work. In our experience, CAaaS lowers the barriers and risk to organizational change, fosters innovation and experimentation, and provides the agility required to meet our customers' increasing and changing needs.